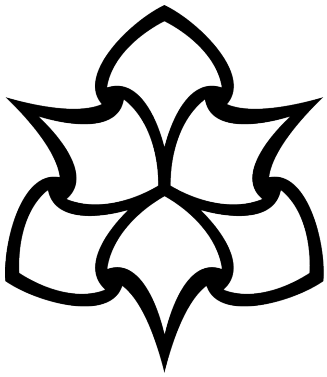

Responsiveness of Single versus Composite Measures of Pain in Knee Osteoarthritis

Matthew J Parkes[1,2], Michael J Callaghan[1,2,3,4], Leslie Tive[5], Mark Lunt[1,2], David T Felson[1,2,6]

[1]Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK.

[2]NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

[3]Faculty of Health, Psychology, and Social Care, Department of Health Professions, Manchester Metropolitan University, Manchester, UK

[4]Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

[5]Pfizer Inc., New York, NY, USA

[6]Clinical Epidemiology Unit, Boston University School of Medicine, Boston, MA, USA.

Address correspondence:

Matthew Parkes, Research Statistician, Research in Osteoarthritis Manchester (ROAM), Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UNITED KINGDOM, M13 9PT (matthew.parkes@manchester.ac.uk)

Word counts:

Abstract: 243

Main Text: 3,495

ORCIDs of authors, where available:

Matthew Parkes, ORCID: 0000-0002-1574-9933

Michael Callaghan, ORCID: 0000-0003-3540-2838

David Felson, ORCID: 0000-0002-2668-2447

Mark Lunt, ORCID: 0000-0002-2391-5575

Data Sharing and Integrity

The corresponding author (MJP) had full access to the data presented in this study, and takes responsibility for the integrity of the data, and the accuracy of the data analysis.

Abstract

Objective

In rheumatoid arthritis, composite outcomes constructed from a combination of outcome measures are widely used to enhance responsiveness (sensitivity to change) and comprehensively summarize response. WOMAC pain is the primary outcome measure in many osteoarthritis (OA) trials. Information from other outcomes, such as rescue medication use, and other WOMAC subscales, could be added to create composite outcomes, but the sensitivity of such a composite has not been tested.

Method

We used data from a completed trial of Tanezumab for knee OA (NCT00733902). The WOMAC questionnaire and rescue medication use were measured at multiple time points, up to 16 weeks. Pain and rescue medication outcomes were standardised and combined into 3 composite outcomes via principal components analysis to produce one score (composite outcome) and their responsiveness was compared to WOMAC pain, the standard. We pooled all treatment doses of Tanezumab into one 'treatment' group, for simplicity, and compared this to the control group (placebo).

Results

The composite outcomes showed modestly but not statistically significantly greater responsiveness when compared to WOMAC pain alone. Adding information on rescue medication to the composite improved responsiveness. While improvements in sensitivity were modest, the required sample sizes for trials using composites was 20-40% less than trials using WOMAC pain alone

Conclusion

Combining information from related, but distinct, outcomes considered relevant to a particular treatment improved responsiveness, could reduce sample size requirements in OA trials and might offer a way to better detect treatment efficacy in OA trials.

Keywords: Osteoarthritis; Outcomes; Pain; Sensitivity to Change; Responsiveness.

Running headline: Sensitivity to Change of Pain Oucomes

Significance & Innovations:

- This study attempts to evaluate meaningful ways of combining single outcomes in a way that improves responsiveness, gaining more power to detect treatment effects without collecting more data.
- This can improve efficiency in future clinical trials, as it helps improve detection of smaller treatment effects with fewer participants.
- Combining outcomes appears to produce composites with greater sensitivity to change than constituent parts.

Introduction

Clinical trialists have a tendency to measure many outcomes. Several of these outcomes (often deliberately) cover overlapping 'domains', attempting to ensure that the 'signal' of a true change in an outcome following an intervention is captured. Pain is a good example; researchers will often use a variety of similar pain-related outcomes in interventional trials.

Pain is a complex, multidimensional measure(1,2), and generating just one scale or item that adequately captures most, if not all, aspects of pain is challenging. Furthermore, as pain is strongly related to functional limitation(3), the most appropriate pain outcome might cover aspects of both pain and function. The optimal clinical trial pain outcome(s) should additionally be sensitive to change following an intervention, by which we mean the outcome's ability to detect a change, often also termed an outcome's responsiveness(4), discriminating well between a true signal, (treatment effect) and noise (random variation).

Composite outcomes are a way of combining (often related) indices or scores to form one overall outcome. This approach, which has been used in many disease areas, including osteoarthritis(5), rheumatoid arthritis(6–8) and asthma(9), may improve the capture of a domain more completely as it takes account of more information than one outcome alone. Pain measurement appears particularly suited to this approach, given its complexity. Combining information from several different domains may improve a composite's ability to detect a change when one truly occurs, and therefore 'responsiveness' may also be improved.

ConstructingComposites:AvailableMethods

There are several methods for combining outcomes into composites. Some of these facilitate domain coverage; others increase responsiveness. Ideally, the method used should improve both. The simplest

method of combining two or more outcomes is through summing or averaging them(5). This method assumes that the constituent outcomes have equal weighting in the composite, and that units from the constituent outcomes are comparable and exchangeable.

A second method of combining multiple outcomes is through the use of weights to assign 'importance' of constituent outcomes. The composite is produced my multiplying each constituent outcome by its weight, and then summing these scores. An example of this is the DAS28(6,7). Weights can be derived from a variety of sources, including statistical modelling (as with the DAS), but also from group consultation, for example via a Delphi exercise (10–14).

Another data-driven approach uses principal components analysis, a data-reduction method which inherently concentrates as much variance from constituent outcomes into as few factors as possible. This method may produce a composite outcome which more completely captures the variance from an underlying multidimensional process, such as pain.

Theoretically, combining several outcomes purporting to measure aspects of pain and its consequences, such as function loss and rescue medication use, should increase domain coverage (as each outcome contributes some information about the pain signal), and therefore responsiveness. Since all of the contributing outcomes should measure that same latent factor (pain), the analysis model used should assume a priori a one factor solution, rather than generating multiple outcomes. This way, we can combine all outcomes related to pain into one composite outcome, which will hopefully show maximal responsiveness in pain.

This study sought to combine several pain outcomes using principal components analysis, taken from a large completed clinical trial of a treatment that reduced pain, and compare the relative responsiveness of these composites to the uncombined WOMAC pain subscale score alone, to establish whether the

inclusion of additional pain information improves responsiveness following administration of an intervention.

Assessment of responsiveness is optimal in certain trial designs. The ideal trial should contain a treatment arm with an intervention which is known to truly change the construct of interest (pain, for example); a control arm which is known to truly not change the construct of interest, and at least two (ideally more) time points in both arms, over which the change in each outcome is assessed. The trial we selected had these features. If the outcome of interest is not changed during the study, then it is not possible to assess responsiveness.

Methods

The data used in this study were from a large completed clinical trial of Tanezumab in participants with knee osteoarthritis (NCT00733902). This trial was a 32 week four-arm parallel-group phase III trial, comparing 3 doses of tanezumab (2.5, 5, or 10 mg) against placebo. Participants were observed at baseline, 2, 4, 8, 12, 16, 24, and 32 weeks; we used data from the 2 week visit to the 16 week visit, as data for rescue medication use were collected only at these visits. For simplicity, we pooled all tanezumab doses (2.5mg, 5mg, and 10mg) together into one 'treatment' group and compared this to the placebo group. Further details regarding the trial's design as well as data on unstandardised outcome scores in have been published previously(15). This study is a reanalysis of completed clinical trial data, and is exempt from ethical review under the NHS Health Research Authority Guidelines.

VariableDefinitions

Single Outcomes:

We used the following pain-related outcomes featured in NCT00733902: WOMAC pain, stiffness, and function subscales; and number of rescue medication pills taken per week.

Composite Outcomes:

Including information from at least two, and up to four outcomes in each composite gives 11 possible combinations available from which composites could be generated. We generated a total of three composite outcomes which were felt to be the most meaningful and pragmatic of the 11 possible combinations. Composite one consisted of the WOMAC pain subscale plus rescue medication. Composite two consisted of all three WOMAC subscales (pain, stiffness, and function). Composite three consisted of all three WOMAC subscales, plus the rescue medication outcome. Composite outcomes were derived by

including the selected combination of variables in a principal components analysis (PCA), which assumed a one factor solution. We opted for PCA, given its propensity to maximise the amount of variance captured in the first (and in this case, only) derived component. We assumed that all included outcomes measured different aspects of one latent (multidimensional) pain variable, and forcing a one component solution therefore ensured that this variable was derived. Angst et al. (2005) found that unrestricted factor analysis of individual WOMAC items established new factors which drew from both the pain and function subscales, and merged them together(16), supporting this idea. It also simplifies the analysis, as it creates only one composite outcome, rather than allowing many composite factors to be generated in each PCA model. We constructed three PCA models, each generating one of three composite outcomes. Rotation of the factor solution (of any type, varimax, promax, or other) was not indicated in our approach, as a one factor solution has only one possible orientation.

We pooled together data from all study visits in the analysis models (rather than using data from baseline only, for example) assuming that it was best to include the maximum available number of observations in the PCA models.

AnalysisApproach

All composite outcome measures were compared to the WOMAC pain, assuming this as the standard.

All of the single outcomes (WOMAC pain subscale score, WOMAC function subscale score, WOMAC stiffness subscale score, and number of rescue medication pills taken) were standardised prior to inclusion in the factor analysis models, and the composites (composites one, two, and three detailed above) were also standardised. Having all variables standardised (as z-scores) allows direct comparison

of outcomes with different units.

We used a random-effects panel linear regression (via SAS's PROC MIXED) to assess change in the standardised outcome score over time, with outcome type, study visit, and treatment group (either tanezumab or placebo) and all possible interactions, as predictor variables. Constructing the data in 'long format', and using outcome type as a categorical dummy-coded variable allows direct testing for differences in responsiveness between all outcomes in one statistical model (further detail on model terms in online appendix). SAS's PROC MIXED command uses a likelihood-based approach, treating missing observations as missing-at-random.

We used linear combinations of coefficients from the regression model (using SAS PROC ESTIMATE) to produce the difference in standardised change between the WOMAC pain subscale and each composite outcome, at each study time point. This formally tests whether the outcomes differed from the WOMAC pain subscale in terms of responsiveness at each of the five time points in the study.

Statistical analysis used SAS® software version 9.3; (SAS Institute Inc., Cary, NC, USA). A nominal alpha level of 0.05 was used for all confidence intervals.

Results

StudySampleDemographics

At baseline, the placebo group (N=172) comprised 119 females (69.2%), with a mean age of 62.2 years, Kellgren-Lawrence grades 2, 3 and 4 of 39.5%, 47.7%, and 12.8% respectively, mean WOMAC pain subscale score (0-10) of 7.1, and mean WOMAC function subscale score (0-10) of 6.6. The pooled tanezumab group (N=518) at baseline had 301 females (58.1%), with a mean age of 61.4 years, Kellgren-Lawrence grades 2, 3 and 4 of 38.4%, 46.3%, and 14.5% respectively, mean WOMAC pain subscale score (0-10) of 7.1, and mean WOMAC function subscale score (0-10) of 6.8.

Ten participants had missing observations for all outcomes at the time points of interest, giving a total sample size for this analysis of 680, compared with the original trial sample size of 690, with 509 in the pooled tanezumab group, and 171 in the placebo group. Data for the 680 included patients could have been collected on 7 outcomes, at 5 time points, giving a total of 23,800 possible observations. Of these, 20,597 were actual observed data points, with 3,203 observations missing (13.5%).

PrincipalComponentsAnalysisResults

The PCA process generated composites with component loadings shown in table 1. WOMAC pain and stiffness subscales consistently had the greatest, and indeed equal, loading, closely followed by the

WOMAC function subscale. When all 3 WOMAC subscale variables were included in the PCA model (in composite 3), the rescue medication's loading dropped considerably.

CompositeOutcomePerformance

All composites showed responsiveness greater than at least some of their constituent outcomes on their own, and this difference was consistent across multiple time points (figure 1). Composite one showed

consistently greater responsiveness than the WOMAC pain subscale alone. The remaining two composites displayed responsiveness greater than all other constituent outcomes, except the WOMAC stiffness subscale. None of the single or composite outcomes showed consistently statistically significantly better responsiveness than that observed in the WOMAC pain subscale at the chosen alpha level (table 3).

We next examined the impact of the observed differences in responsiveness on sample size requirements for a hypothetical new trial featuring the same design (table 2). For example, the WOMAC pain subscale between-groups standardised change at four weeks was a difference of -0.37. A hypothetical new trial of identical design observing this between-group difference for the WOMAC pain outcome would require 236 participants (118 per group) to achieve 80% power with a two sided 5% type-I error rate. In contrast, using composite 1 (i.e. including information on rescue medication as well as the WOMAC pain subscale score) as the primary outcome which had an observed difference at four weeks of -0.41, the same trial would need 190 participants (95 per group) to achieve 80% power with this difference - a saving of 46 participants. When the observed differences between treatments is smaller, the reduction in sample size was more extreme: the WOMAC pain difference at 16 weeks (-0.26) would require 476 participants for 80% power in a hypothetical new trial, compared to only 364 participants when using

162

composite 1 (using the observed difference of 0.29), a saving of 112 participants.

Discussion

We found that composite outcomes generally had moderately greater responsiveness in a large OA trial than WOMAC pain – the usual standard outcome of these trials. That suggests if one of these composite outcomes were used as the primary outcome in an OA trial, fewer subjects would be needed to demonstrate treatment efficacy.

The improvements in responsiveness did not meet the criteria for a statistically significant difference, but perhaps a more salient measure of their import was to determine what effect using these outcomes had on the sample size needed to be likely to show statistically significant effects of treatment vs. placebo. We found that the reduction in sample size was substantial, ranging from roughly 20 to 40%. Thus, composites could substantially diminish the sample sizes needed in an osteoarthritis trial whose main outcome is pain.

Eigenvalues from the three composite models all were much greater than the 1.0 cut-off typically used to select retained factors in a PCA model(17), and a large proportion of the variance in the outcomes was captured by the first component in the PCA model, as anticipated (table 1). The second factor listed in the model output (which was not extracted in this analysis) in all cases had an eigenvalue much less than 1,

lending support to the idea that the selected correlated outcomes are well captured in in one 'pain' component.

Rescue medication use, whilst contributing to the 'pain' component the least (table 1), appeared to still improve responsiveness: composites including this outcome - composites one (WOMAC pain plus rescue medication use) and three (WOMAC pain, stiffness, and function, plus rescue medication) - showed slight improvements in responsiveness compared with composite 2, which excluded rescue medication.

Aside from the methods used to combine outcomes, the method chosen to assess responsiveness is also important(18,19). Several methods are commonly cited to quantify responsiveness: the standardised response mean (SRM)(20), the effect size (ES)(18), either Glass' Δ(21) or Cohen's d(22) (depending on the standard deviation used), or Guyatt's responsiveness index (GRI) (23). All of these methods have two important limitations. First, all methods calculate responsiveness over two time points, and cannot easily be generalised to a study which has three or more time points. This prevents assessment of how responsiveness may fluctuate over time, and limits the definition of responsiveness only to the magnitude of change relative to its variance, rather than the speed of response. Second, these methods do not directly assess statistical inference; a; differences in responsiveness coefficients are assessed descriptively only. Methods have been proposed (modified jackknife procedure (5,24,25), bootstrapping (26)) to address this issue, but other methods which directly perform statistical inference as part of the method generating the coefficient are desirable.

Our approach made use of z-scores (standard scores) (27). Converting each outcome's absolute score to a z-score allows direct comparison of change in an outcome at different time points, thereby allowing direct assessment of change over time, and direct comparison between different outcomes. This

methodology has been used previously to compare non-composite outcomes(28).

The PCA approach assumes that an intervention will alter several related aspects of a common construct, therefore combining all the multidimensional aspects of pain together to form one outcome should increase responsiveness. However, if one aspect of pain is changed alone, then the inclusion of other aspects of pain which do not change may decrease the sensitivity of the composite. Our finding that the

WOMAC stiffness subscale was the most sensitive outcome may fit this explanation: It may be, at least in this trial, that the WOMAC stiffness subscale was the closest correlate to the actual latent pain factor altered by the treatment, hence the greatest responsiveness, and inclusion of other subscales or rescue medication eroded it. Our finding may be limited to tanezumab alone – as the agent's anti-nerve growth factor effect may have a greater impact on the stiffness sensation than other pain subscales(15,29).

Freemantle et al. (2003) provide a comprehensive discussion on the use of composite outcomes in clinical trials(30), highlighting how composite outcomes can obfuscate changes in constituent outcomes. This is particularly problematic when outcomes are unrelated (for example, a composite which combines cardiovascular events and mortality), although they note the statistical advantages (increased power and sensitivity) that arise through the construction of composites(30,31). This discussion highlights how both the outcomes used in the composite, and the method by which they are combined, are important. The present study combined the three WOMAC subscales, pain, stiffness, and function, into one composite outcome. We assumed that these three subscales were all aspects of the same construct (pain). The PCA (table 1) produced extremely high factor loading in all three subscales, supporting this notion - at least in this trial. In contrast, if pain and function were discrete constructs, then the PCA should fail, with either pain or function alone loading on the latent factor. Both Ryser et al. (1999), and Angst et al. (2005) found

close association between pain and function WOMAC subscales, partly supporting this finding(16,32). In addition, an item overlap analysis on the WOMAC pain and function subscales by Stratford et al. (33) found significant item redundancy between the pain and function subscales, and a further factor analysis on the WOMAC items found clustering of items not by subscale, but by activity(34), suggesting that the WOMAC subscales are not distinct.

We surmised that responsiveness in the outcomes may differ over time, as well as in magnitude. In this study, all outcomes appeared to have responded at the same time point, and retained their relative positions consistently over time (none of the outcome's trajectories crossed over each other over time, figure 1).

There are limitations to this analysis. We observed only very few statistically significant differences between outcomes. The trial was designed to observe a difference in the primary outcome between treatment groups (a relatively large difference), and was not designed to compare treatment differences between outcomes (much smaller differences). Therefore even the large sample size in the trial provides relatively low power to detect differences between outcomes. Ideally in future, this analysis would be designed in to the trial prior to commencement, with appropriate sample size and power. We also allowed many interaction effects, which increased model-to-data fit at the expense of statistical power. We have assumed in this analysis that the covariate structure of the pain outcomes, and the relationship between the outcomes and the latent (unobserved) pain outcome are consistent between studies, and therefore generalisable across other studies. This is a relatively strong assumption, requiring validation in other datasets to allow wider generalisation to other trials with confidence.

While the aim of this approach was to include additional information on pain from rescue medication data, this outcome may not be optimal. Rescue medication is a challenging variable to collect data on accurately, and therefore the likelihood is that measurement error of this variable is high. This may provide an explanation for why the improvement in sensitivity of composites including rescue medication are small.

Even though the between-outcome differences were not statistically significant, even a small improvement in responsiveness can impact upon sample size calculations (table 3). This produces gains in efficiency without collecting any novel data simply by reanalysing the data using a method which produces a more sensitive, and therefore efficient, outcome. We could have included further assessment of other composites made from different combinations of the 11 possible from the four single outcomes used, for example one using WOMAC pain plus WOMAC stiffness. We opted to create the three composites which would have the most pragmatic impact on outcome inclusion/exclusion when designing a trial. The alternative, generating all 11 possible combinations and comparing them head to head, would further reduce the statistical power to discern differences between composite outcomes.

The PCA approach to generating a composite outcome by its nature produces a unitless score. While the generated score may have increased responsiveness compared to one of the constituent outcomes, it is more difficult to ascertain the clinical importance of the observed effect, in comparison to another outcome with meaningful units and an agreed minimally clinical importance difference (MCID). A downside PCA composites is the absence of known values of MCID, but this could be established if a specific composite were widely used.

The choice of primary and secondary outcomes in this trial limited the choice of outcomes available to combine into a composite. Ideally, we would have preferred to use a trial featuring a wider range of pain outcomes, particularly the more recent KOOS(35) and ICOAP(36) questionnaires; however a dataset using these outcomes amongst others, and featuring the other requirements was not available.

The present findings are similar to our previous paper, which used data from two other completed clinical trials of non-drug interventions(28). In both of these trials, the WOMAC stiffness subscale also showed an increased, but non-statistically-significant, degree of responsiveness compared to the other two WOMAC subscales. Angst et al. (2001; 2008), in contrast, found the WOMAC pain subscale to be the most sensitive outcome to change(5,24), however these studies did not examine rescue medication, and used a two-time point approach only. Further, the two studies previously analysed were both prospective cohort studies lacking a control group. Thus, optimizing the detection of treatment effect over placebo was not possible in the two Angst et al. analyses.

In summary, we investigated whether collapsing several measures of a multidimensional construct into one composite outcome through the use of PCA could help improve responsiveness following an intervention. Adding rescue medication alongside other elements of the WOMAC showed improved responsiveness, greater than the constituent outcomes.

Acknowledgements

We would like to acknowledge the contributions of Pfizer in allowing our team access to the completed trial datasets, and their support in using their trial analysis platform, specifically Pamela Singletary, Daireen Garcia, Glenn Pixton, and Michael Smith

for their help and support. We would also like to acknowledge the contributions of the ROAM team to this project. The ROAM group is supported by the Manchester Academic Health Sciences Centre (MAHSC). Arthritis Research UK also continues to support the Centre for Epidemiology (grant number 20380). This report includes independent research supported by (or funded by) the National Institute for Health Research Biomedical Research Unit Funding Scheme. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. The funding agencies had no role in any of the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Prof. Felson is supported by NIH AR4778. The authors would also like to acknowledge

the assistance given by Séamus Byers, Contracts, IT Services and the use of the Computational Shared Facility at The University of Manchester.

Contributions of authors

DTF initially proposed the study.

Wrote the manuscript: MJP, DTF

Analysis and interpretation of data: MJP, ML

Reviewed drafts of the paper: MJP, MJC, ML, LT, DTF

Competing Interest Statement (Financial Support)

MJP, MJC, ML, and DTF receive salary support from the National Institute for Health Research, as part of the Manchester Musculoskeletal NIHR Biomedical Research Unit Grant.

LT owns stock or stock options in Pfizer.

References

1. Williams AC de C, Craig KD. Updating the definition of pain. Pain [Internet] 2016;157:2420-3. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00006396-201611000-00006

2. Mease PJ, Hanna S, Frakes EP, Altman RD. Pain mechanisms in osteoarthritis: Understanding the role of central pain and current approaches to its treatment. J Rheumatol 2011;38:1546-51.

3. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthr Cartil 2013;21:1145-53.

4. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005. p. 9-19.

5. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis [Internet] 2001;60:834-40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11502609\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1753825

6. van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49:916-20.

7. van der Heijde DM, van't Hof MA, van Riel PL, van Leeuwen MA, van Rijswijk MH, van de Putte LB. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. Ann Rheum Dis [Internet] 1992;51:177-81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1550400\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1005654

8. Ibrahim F, Tom BDM, Scott DL, Prevost AT. A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. Trials [Internet] Trials; 2016;17:272. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27255212\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4890523

9. Cloutier MM, Schatz M, Castro M, Clark N, Kelly HW, Mangione-Smith R, et al. Asthma outcomes: Composite scores of asthma control. J Allergy Clin Immunol [Internet] Elsevier Ltd; 2012;129:S24-33. Available from: http://dx.doi.org/10.1016/j.jaci.2011.12.980

10. Gossec L, Paternotte S, Aanerud GJ, Balanescu a, Boumpas DT, Carmona L, et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. Ann Rheum Dis 2011;70:935-42.

11. Dechartres A, Albaladejo P, Mantz J, Samama CM, Collet JP, Steg PG, et al. Delphi-consensus weights for ischemic and bleeding events to be included in a composite outcome for RCTs in thrombosis prevention. PLoS One 2011;6:10-2.

12. Rogozinska E, D'Amico MI, Khan KS, Cecatti JG, Teede H, Yeo S, et al. Development of composite outcomes for individual patient data (IPD) meta-analysis on the effects of diet and lifestyle in pregnancy: A Delphi survey. BJOG An Int J Obstet Gynaecol 2016;123:190-8.

13. Monchaud C, Marin B, Estenne M, Preux P-M, Marquet P. Consensus conference on a composite endpoint for clinical trials on immunosuppressive drugs in lung transplantation. Transplantation 2014;98:1331-8.

14. Tong BC, Huber JC, Ascheim DD, Puskas JD, Jr BF, Blackstone EH, et al. Evidence for the Heart Team. 2013;94:1908-13.

15. Brown MT, Murphy FT, Radin DM, Davignon I, Smith MD, West CR. Tanezumab reduces osteoarthritic knee pain: results of a randomized, double-blind, placebo-controlled phase III trial. J Pain [Internet] Elsevier Ltd; 2012;13:790-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22784777

16. Angst F, Ewert T, Lehmann S, Aeschlimann A, Stucki G. The factor subdimensions of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) help to specify hip and knee osteoarthritis. A prospective evaluation and validation study. J Rheumatol 2005;32:1324-30.

17. Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas [Internet] 1960;20:141-51. Available from: http://www.garfield.library.upenn.edu/classics1986/A1986E107600001.pdf

18. Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. Health Qual Life Outcomes 2005;3:23.

19. Norman GR, Wyrwich KW, Patrick DL. The mathematical relationship among different forms of responsiveness coefficients. Qual Life Res 2007;16:815-22.

20. Liang MH, Fossel a H, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.

21. Hedges L V., Olkin I. Statistical methods for meta-analysis. Phytochemistry [Internet] 1985;72:369. Available from: http://oldjll.sustainabilityforhealth.org/trial_records/20th_Century/1980s/hedges/hedges-kp.pdf

22. Cohen J. Statistical power analysis for the behavioral sciences [Internet]. Statistical Power Analysis for the Behavioral Sciences 1988. p. 567. Available from: http://books.google.com/books?id=Tl0N2lRAO9oC&pgis=1

23. Guyatt G, Walter S, Norman G. Measuring Change Over Time- Aseessing the Usefulness of Evaluative Instruments. J Chronic Dis [Internet] 1987;40:171-8. Available from: <Go to ISI>://WOS:A1987G268000010

24. Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med Res Methodol [Internet] 2008;8:26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18439285\nhttp://www.biomedcentral.com/content/pdf/1471-2288-8-26.pdf

25. Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer H-K, Simmen BR. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. Arthritis Rheum [Internet] 2008;59:391-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18311752

391

2 .Spadoni GF, Stratford PW, Solomon PE, Wishart LR. The Evaluation of Change in Pain Intensity: A Comparison of
6 the P4 and Single-Item Numeric Pain Rating Scales. J Orthop Sport Phys Ther 2004;34:187-93.

392 27. Kirkwood BB, Sterne J. Essential medical statistics [Internet]. Malden, MA: Blackwell Science 2003. 1-512 p.
393 Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Essential+Medical+Statistics#0

394 28. Parkes MJ, Callaghan MJ, O'Neill TW, Forsythe LM, Lunt M, Felson DT. Sensitivity to Change of Patient-Preference
395 Measures for Pain in Patients With Knee Osteoarthritis: Data From Two Trials. Arthritis Care Res 2016;68:1224-
396 31.

397 29. Lane NE, Schnitzer TJ, Birbara C a, Mokhtarani M, Shelton DL, Smith MD, et al. Tanezumab for the treatment of
398 pain from osteoarthritis of the knee. N Engl J Med 2010;363:1521-31.

30. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003;289:2554-9.

31. Freemantle N, Calvert MJ. Interpreting composite outcomes in trials. Br Med J 2010;341:c3529.

32. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. Arthritis Care Res [Internet] 1999;12:331-5. Available from: <Go to ISI>://WOS:000083009200004\nhttp://onlinelibrary.wiley.com/store/10.1002/1529-0131(199910)12:5<331::AID-ART4>3.0.CO;2-W/asset/4_ftp.pdf?v=1&t=i8geidtl&s=8752fa134a34dd764789f1c25066301b7320e6e6

33. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes [Internet] 2003;1:64. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=280702&tool=pmcentrez&rendertype=abstract

34. Hawker GA, Davis AM, French MR, Cibere J, Jordan JM, March L, et al. Development and preliminary psychometric testing of a new OA pain measure - an OARSI/OMERACT initiative. Osteoarthr Cartil 2008;16:409-14.

Figure Legend

Figure 1: Sensitivity to Change of Single Pain-Related Outcomes from the Tanezumab Trial. Values plotted are the control-treatment differences in standard score, at different study time points. More negative values indicate increased sensitivity to change